IJSPT

ORIGINAL RESEARCH

# INTERRATER AND TEST-RETEST RELIABILITY OF THE Y BALANCE TEST IN HEALTHY, EARLY ADOLESCENT FEMALE ATHLETES

**Eric T. Greenberg, PT, DPT, SCS[1]**
**Matthew Barle, PT, DPT[1]**
**Erica Glassmann, PT, DPT[1]**
**Min-Kyung Jung, PhD[2]**

## ABSTRACT

***Background:*** Adolescence is the stage of development marked by peak rates of skeletal growth resulting in impaired dynamic postural control and increased injury risk, especially in female athletes. Reliable tests of dynamic postural control are needed to help identify athletes with balance deficits and assess changes in limb function after injury.

***Purpose:*** To estimate the interrater and test-retest (intrarater) reliability of the Y-Balance Test in a group of early adolescent females over a one-month period when administered by novice raters.

***Methods:*** Twenty-five early adolescent females (mean age 12.7 $\pm$ 0.6 years) participated. Two physical therapy student raters, randomly selected from a pool of five, simultaneously assessed each subject's performance on the Y-Balance Test and were blinded to each other's results. Twenty-one subjects returned for a second session (mean 32.3 $\pm$ 9.6 days) and were assessed by the same two raters, blinded to previous measurements. Maximum and normalized reach distances and composite scores of the right and left limbs were collected. Intraclass correlation coefficients (ICC) were calculated for between rater and between session agreement. Measurement error and minimal detectable change values were calculated for clinical interpretation.

***Results:*** Interrater reliability was excellent for all reach directions and composite scores of the right limb (ICC 0.973-0.998) and left limb (ICC 0.960-0.999) except for the day 1 left anterior reach which was good (ICC 0.811). Test-retest reliability were moderate to excellent for the right limb (ICC 0.681- 0.908) and moderate to good for left limb (ICC 0.714 - 0.811). Minimal detectable change values for the right and left limbs ranged between 2.02-3.62% and 2.77-3.63%, respectively.

***Conclusions:*** The Y-Balance Test is a reliable tool to assess dynamic balance in early adolescent females and may be utilized in a clinical setting to monitor function over a one-month time interval. Between rater differences were mainly attributed to disparities in subjective test requirements and not quantitative measures of reach distance.

***Level of Evidence:*** Level 2

***Key Terms:*** adolescent female, dynamic balance, movement system, reliability, Y-Balance Test

[1] Department of Physical Therapy, New York Institute of Technology, Old Westbury, NY
[2] College of Osteopathic Medicine, New York Institute of Technology, Old Westbury, NY

**Conflict of Interest Statement:** We, the authors, affirm that we have no financial or commercial affiliations related to the performance or outcome of this manuscript.

**Grant Support:** There was no grant support for this study.

**IRB Statement:** Institutional Review Board of the New York Institute of Technology approved this study. Protocol number: BHS-1275

**CORRESPONDING AUTHOR**
Eric T. Greenberg, PT, DPT, SCS
New York Institute of Technology
Northern Blvd, P.O. Box 8000
Department of Physical Therapy
Old Westbury, NY
E-mail: egreen05@nyit.edu
Telephone: 516-686-7912
Fax: 516-686-7699

## INTRODUCTION

Adolescence is the stage of development characterized by accelerated rates of physical growth,[1] redistribution of adipose tissue,[2] increased length of long bones,[3] and increased joint forces and torques.[4] To better attempt to homogenize the wide range of physical, emotional, and cognitive changes that occur during this phase, adolescence is dichotomized into two distinct phases. Though widely variable, early adolescence, typically begins around age 11 in females and 14 in males while late adolescence begins a few years later, and may persist into the third decade of life.[5] Despite the benefits of youth sports participation,[6] adolescent females are at an increased risk of lower extremity injuries in comparison to their male counterparts, [4,7-12] and may be partially explained by the absence of a neuromuscular spurt and resultant muscle strength and recruitment pattern deficits.[13,14] Additionally the incidence of ACL injuries are greatest during the high school years[15] and recommendations support the implementation of targeted neuromuscular control interventions to high risk populations prior to the time of peak injury risk.[16]

Sophisticated measures of dynamic postural control, such as stabilometry, are able to detect subtle deficits in young athletes,[17] but are expensive and may not be readily available in a clinical setting. The Y-Balance Test (YBT) is a low-cost, clinical measure of dynamic balance that mimics the demands of sports requiring unilateral balance. The YBT assesses limb symmetry utilizing a unilateral lower extremity reaching task in three different directions (anterior [ANT], posteromedial [PM], and posterolateral [PL]) (Figure 1). When used as a screening tool, an anterior reach asymmetry greater than 4cm between limbs has been associated with an increased risk of sustaining a lower extremity injury in division I collegiate athletes.[18] It has also been used as a clinical outcome measure to gauge functional improvement and guide activity progression following injury.[19-21] The reliability of the YBT has been studied on various populations[22-24] by raters of varying levels of experience.[25] Plisky et al[23] reported good to excellent interrater (ICC 0.99-1.0) and intrarater (ICC 0.85-0.91, 95% CI 0.62 -0.96) reliability in individual reach directions and composite scores when assessed by experienced raters in a group of healthy young adults with a mean age of 19.7 years old. Similar findings of interrater (ICC 0.85 to 0.93, 95% CI 0.75-0.96) and intrarater (ICC 0.80-0.85 , 95% CI 0.68-0.91) reliability have been reported when the YBT was administered by raters with minimal testing experience.[25] Additionally, Faigenbaum et al[24] reported excellent interrater reliability (ICC >0.995) and moderate to good between session intrarater reliability ($0.907 \leq ICC \leq 0.974$) in a group of preadolescents ages 6-12 years old.

Though the YBT has been found to be a reliable tool for preadolescent, late adolescent, and adult athletes, developmental differences exist between

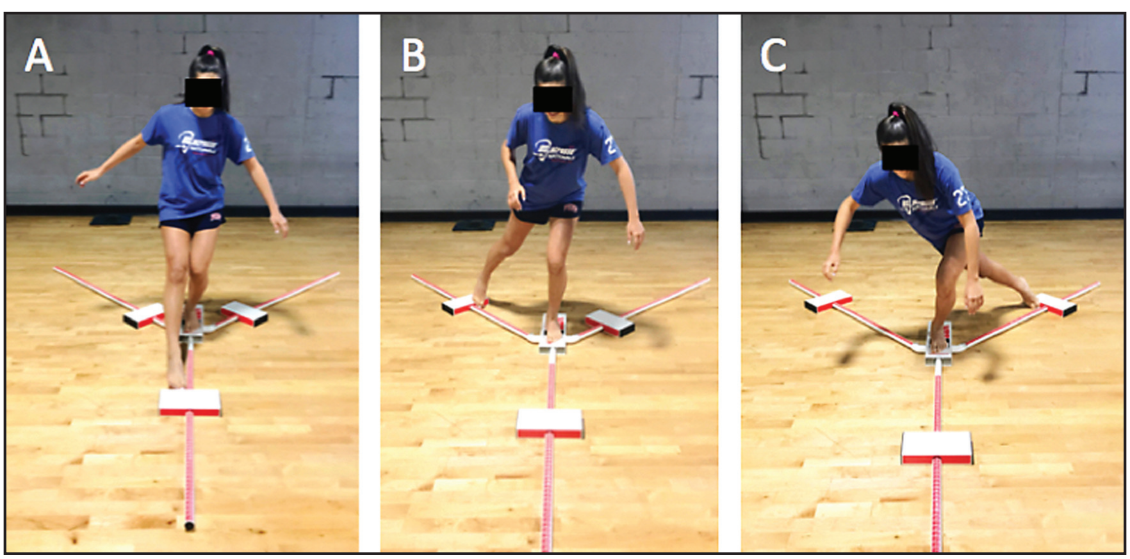


**Figure 1.** *Performance of Y Balance Test (A) left anterior reach, (B) left posteromedial reach, and (C) left posterolateral reach.*

these populations and the early adolescent female. It is also unknown whether the YBT may be used for pre-participation screening purposes within this population by those with limited experience in assessing human movement such as coaches, physical education teachers, and students. Additionally, due to the rapid growth during this maturational phase, outcome measures should be stable over typical, clinical test-retest time intervals to ensure improvements in test performance actually reflect true functional change. Therefore, the purpose of this study was to assess the inter- and intra- rater (test- retest) reliability of the YBT in a group of early adolescents ages 12-14 and over a one month period when administered by novice practitioners. It was hypothesized that YBT is a reliable tool to assess the dynamic postural control of early adolescent female athletes when administered by novice raters.

## METHODS

### Participants

A convenience sample of 26 multisport female athletes, ages 12-14, was recruited from local community-based recreational programs between May and September of 2017. Participants had no prior experience performing the YBT. To be included in the study, participants demonstrated > 35 degrees of ankle range of motion on the dorsiflexion lunge test, with no more than a 5-degree side-to-side difference, and ability to stand unsupported on one leg for at least five seconds without a loss of balance. Exclusion criteria included lower extremity amputation, cognitive deficits, vestibular disorders, blindness in at least one eye, current or undergoing treatment for inner ear, respiratory infection, or head cold, cerebral concussion within the prior six months, lower extremity injury in the prior three months (diagnosed by a medical professional and missed day of athletic or recreational activity), or lack of medical clearance for athletic participation. This study was approved by the New York Institute of Technology Institutional Review Board and written informed consent was obtained from parents/legal guardians along with subject verbal and signed assent prior to study participation.

### Study Protocol

Participants were asked to perform the YBT on two separate days, three to five weeks apart. After screening for inclusion and exclusion criteria, each subject viewed a standardized video recording outlining the YBT testing procedure. Any questions regarding YBT performance were answered at this time. Two raters were randomly selected from a pool of five raters. Raters were all students in their second-year doctor of physical therapy program and were trained to administer the YBT via a one hour web-based tutorial. Raters had an additional two-hour practice session to familiarize themselves with the YBT procedure and instrument. Supervision was provided by a licensed physical therapist with 13 years of clinical experience who utilizes the YBT clinically.

### YBT Protocol

YBT testing protocol was similar to the one previously described by Plisky, et al.[23] All practice and testing was performed on the commercially available Y-Balance Test Kit™ (Move2Perform, Evansville, IN). To perform the YBT, the participant stood barefoot with one foot on the center foot plate and the most distal aspect of the toes just behind the starting line. Reach side and direction was operationalized in reference to the stance limb. While maintaining single-leg stance, the subject was instructed to push the reach indicator with the reach foot as far as possible and return to the original start position. A trial was determined unsuccessful if the subject failed to maintain unilateral stance on the platform, kicked the reach indicator, used the reach indicator for support, or did not return to the start position under control.[23] Reach distance was measured at the nearest edge of the reach indicator to the closest 0.5 centimeter.

Prior to formal testing, each subject performed six practice trials in each of the three reach directions on each leg.[24,26] During practice trials, raters highlighted the importance of maintaining single leg stance and offered feedback regarding errors in test performance. Following the these trials, a rest period was allotted where one rater recorded the subject's height to the nearest 0.5 cm using a wall mounted measuring tape, body mass to the nearest 0.1kg using a digital scale, and leg length in supine to the nearest 0.5cm, measured from the subject's ASIS to the medial malleolus.

Testing order for formal test trials was standardized across all subjects and included three sequential reaches in each of the six directions; right ANT, left ANT, right PM, left PM, right PL and left PL. Each rater was blinded to the other's results and no feedback was offered as they simultaneously observed a single trial. To reduce bias and analyze reasons for interrater discrepancies (quantitative measurement error or subjective decisions regarding test success) each rater recorded the distance reached and if the trial was successful or unsuccessful, but did not share this information with the subject or other rater. After three trials, the raters were asked if they had recorded a least one successful trial for the respective reach direction. If they did not, the subject was asked to perform an additional trial, in that particular direction, until a successful trial was recorded. Three to five weeks later, an identical testing protocol was performed by the same two raters, blinded to the first day results. (Figure 2) This test-retest time frame was chosen to mimic a typical interval between reassessments often utilized in a clinical setting.

The maximum reach distances for each direction and composite (COMP) scores (sums of the maximum distances for each limb) were utilized for analysis. To allow for between subject comparisons, both individual reach and COMP scores were normalized by dividing by the leg length of the contralateral (reaching) limb. Scores were calculated with trials that were deemed successful and adhered to all YBT standards. To examine differences between raters' ability to determine test success, "modified" YBT scores were also calculated. These scores differed from typical YBT protocol as they utilized maximum reach distances from both successful and unsuccessful trials and disregarded raters' subjective decisions pertaining to trial success (i.e. inability to maintain unilateral stance on the platform, kicked the reach indicator, used the reach indicator for support, or did not return to the start position under control).

The Guidelines for reporting reliability and agreement studies (GRRAS) was used to ensure the quality of reporting the findings of this study.[27]

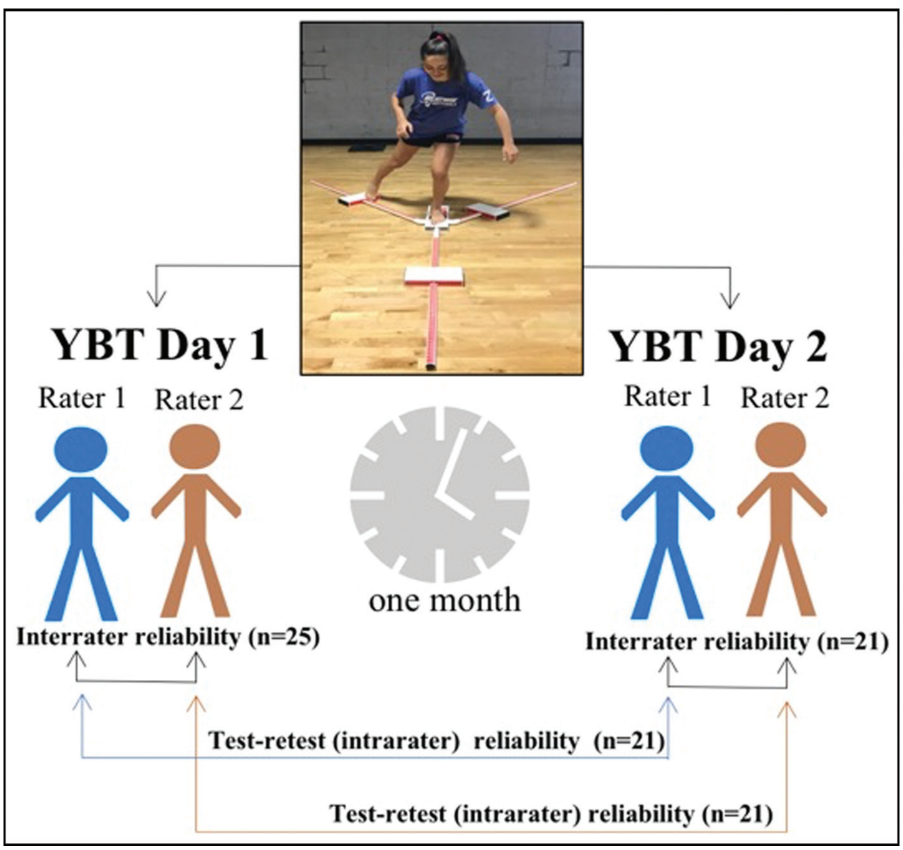


**Figure 2.** *Schematic of Study Protocol.*

### Statistical Analysis

Means and standard deviation or medians and the range as the interval from minimum to maximum were computed to describe demographic data and the outcome measurements. Paired t-tests were applied to test the significance of change between sessions for the demographic variables. Intraclass correlation coefficient (ICC) in an appropriate model was computed to assess the reliability as the agreement between raters (ICC 2, 1) and between sessions (ICC 3, 1). 95% confidence interval of the ICC was also computed to estimate its precision. The range of ICC values was described using the classification described by Fisher,[28] where ICC $<$ 0.5 was considered poor, 0.5$<$ ICC$<$ 0.75 was considered moderate, 0.75$<$ ICC $<$ 0.9 was considered good and ICC$>$ 0.9 is was considered excellent. Measurement error was assessed by the square root of mean squared error (RMSE) and the standard error of measurement (SEM). The minimal detectable change (MDC) computed as $1.96\times\sqrt{2}\times$ SEM was reported for clinical interpretation. All statistical procedures were performed using SPSS version 23 (IBM SPSS Statistics., Armonk, NY).

## RESULTS

Of the twenty-six subjects recruited and screened, one subject was deemed ineligible (did not meet dorsiflexion ROM requirements) for a total of 25 participants. Of those 25 that participated in the first day of testing, two subjects did not return for the second day of testing while two others sustained a lower extremity injury between testing sessions (toe fracture and ankle sprain) for a total of 21 participants who completed both days of testing. Subject characteristics (Table 1) and performance data were analyzed and subgrouped to correct for subjects lost to drop out: first day of testing for all 25 subjects, first day of testing of those subjects that returned for the second day of testing, and those that returned on the second day of testing. There were significant differences in height ($p = 0.016$) and weight ($p = 0.003$) between sessions in the subgroup of 21 subjects that participated in both test days.

YBT mean reach distances, standard deviations, median reach distances, range, and normalized reach distances of each limb and direction are reported in Table 2. Interrater reliability calculations were calculated separately for the 25 pairs of observations from day one and 21 pairs of observations from day 2. Test-retest (intrarater) reliability calculations included the 21 pairs of between session observations made by the same rater, for a total of 42 data points.

### Interrater reliability

Interrater reliability of YBT scores are presented in Table 3. Day 1 values were excellent for the all reach directions and COMP scores of the right limb (ICC 0.973-0.992) and left limb (ICC 0.960-0.989) except for the left ANT reach which was good (ICC 0.811). For day 2, interrater reliability of YBT scores were excellent for all reach directions and COMP scores of the right limb (ICC 0.988-0.998) and left limb (ICC 0.993-0.999).

Within session total error expressed as a percentage of the mean reach distance was greatest for the left ANT direction of day 1 of testing (6.26%) with all other reaches for both days being less than 3% error. SEM was less than 2% of the mean reach distance for all directions and limbs.

**Table 1.** *Subject Demographics.*

| | Age | Height (cm) | Weight (kg) | Right limb length (cm) | Left limb length (cm) |
|---|---|---|---|---|---|
| **Day 1 (n=25)** | 12.7 (0.6) | 154.0 (9.1) | 46.4 (9.6) | 81.6 (5.5) | 81.6 (5.7) |
| **Day 1 (n=21)‡** | 12.7 (0.6) | 154.0 (9.9) | 45.9 (9.8) | 81.7 (5.9) | 81.7 (6.3) |
| **Day 2 (n=21)** | 12.8 (0.6) | 154.9 (9.8) | 47.5 (10.8) | 82.1 (6.4) | 82.1(6.4) |
| **p value between day 1 and 2‡†** | 0.08 | 0.016* | 0.003* | 0.20 | 0.24 |

cm= centimeters; kg= kilograms
‡ Subgroup of subjects from day 1 not lost to follow up
†Average time between test days 1 and 2 = 32.3 days (9.6)
*p< 0.05

**Table 2.** *YBT Reach Distances.*

| | Mean reach in cm (SD) | | Median Reach in cm (range) | | Normalized mean reach† (SD) | |
|---|---|---|---|---|---|---|
| | Right | Left | Right | Left | Right | Left |
| **Day 1 (n=25)** | | | | | | |
| **ANT** | 56.5 (6.2) | 57.5 (5.8) | 57.3 (43.0, 67.3) | 58.0 (48.0, 68.0) | 69.3 (6.7) | 70.5 (5.4) |
| **PM** | 84.4 (9.6) | 85.8 (8.8) | 84.0 (64.8, 106.0) | 85.0 (67.5, 101.8) | 103.5 (10.6) | 105.3 (10.1) |
| **PL** | 83.8 (8.7) | 85.3 (10.6) | 83.3 (63.0, 102.0) | 86.5 (61.5, 105.5) | 102.9 (10.8) | 104.6 (11.4) |
| **COMP** | 224.7 (21.0) | 228.6 (22.8) | 226.8 (179.0, 265.8) | 234.5 (183.5, 269.8) | 91.9 (7.9) | 93.5 (7.9) |
| **Day 1 (n=21)** | | | | | | |
| **ANT** | 57.0 (6.6) | 57.1 (5.9) | 58.0 (43.0, 67.3) | 57.0 (48.0, 68.0) | 69.8 (7.2) | 69.9 (5.2) |
| **PM** | 84.0 (9.7) | 85.0 (8.9) | 84.0 (64.8, 106.0) | 85.0 (67.5, 101.8) | 103.0 (10.5) | 104.1 (10.0) |
| **PL** | 83.4 (8.3) | 85.3 (10.7) | 83.3 (63.0, 98.3) | 86.5 (61.5, 105.5) | 102.3 (10.4) | 104.4 (11.4) |
| **COMP** | 224.4 (21.6) | 227.4 (23.6) | 227.8 (179.0, 265.8) | 234.5 (183.5, 269.8) | 91.7 (8.0) | 92.8 (8.0) |
| **Day 2 (n=21)** | | | | | | |
| **ANT** | 57.8 (6.6) | 57.8 (6.3) | 61.0 (46.5, 67.5) | 59.0 (47.0, 69.5) | 70.6 (7.1) | 70.5 (6.0) |
| **PM** | 85.1 (8.3) | 87.8 (8.9) | 86.0 (71.0, 101.5) | 89.0 (71.8, 101.0) | 104.0 (10.1) | 107.1 (9.0) |
| **PL** | 84.7 (7.5) | 86.4 (8.6) | 84.5 (71.8, 99.0) | 87.0 (70.3, 104.5) | 103.4 (8.6) | 105.5 (9.1) |
| **COMP** | 227.6 (19.0) | 232.0 (21.3) | 230.3 (196.3, 264.0) | 232.5 (193.8, 269.0) | 92.7 (7.1) | 94.4 (6.7) |

cm= centimeters; SD= Standard Deviation; ANT= Anterior; PM= Posteromedial; PL= Posterolateral; COMP= Composite
†normalized to limb length expressed as a percentage

**Table 3.** *YBT reach distance interrater reliability and error.*

| | ICC $_{(2,1)}$ (95% CI) | | Within session total error (% mean reach distance) | | SEM (% mean reach distance) | |
|---|---|---|---|---|---|---|
| | Day 1 | Day 2 | Day 1 | Day 2 | Day 1 | Day 2 |
| **Right Limb** | | | | | | |
| ANT | 0.984 (0.965, 0.993) | 0.995 (0.988, 0.998) | 1.10 (1.95%) | 0.68 (1.17%) | 0.22 (0.39%) | 0.15 (0.26%) |
| PM | 0.992 (0.981, 0.996) | 0.988 (0.971, 0.995) | 1.27 (1.50%) | 1.28 (1.50%) | 0.25 (0.30%) | 0.28 (0.33%) |
| PL | 0.973 (0.940, 0.988) | 0.998 (0.996, 0.999) | 2.05 (2.44%) | 0.44 (0.52%) | 0.41 (0.49%) | 0.10 (0.11%) |
| COMP | 0.987 (0.972, 0.994) | 0.996 (0.990, 0.998) | 3.35 (1.49%) | 1.72 (0.75%) | 0.67 (0.30%) | 0.37 (0.16%) |
| **Left Limb** | | | | | | |
| ANT | 0.811 (0.616, 0.912) | 0.993 (0.982, 0.997) | 3.60 (6.26%) | 0.79 (1.37%) | 0.72 (1.25%) | 0.17 (0.30%) |
| PM | 0.960 (0.911, 0.982) | 0.997 (0.993, 0.999) | 2.57 (3.00%) | 0.69 (0.79%) | 0.51 (0.60%) | 0.15 (0.17%) |
| PL | 0.989 (0.976, 0.995) | 0.999 (0.998, 1.000) | 1.59 (1.86%) | 0.37 (0.43%) | 0.32 (0.37%) | 0.08 (0.09%) |
| COMP | 0.982 (0.960, 0.992) | 0.998 (0.995, 0.999) | 4.37 (1.91%) | 1.36 (0.59%) | 0.87 (0.38%) | 0.30 (0.13%) |

ICC= Intraclass correlation coefficient; CI= Confidence Interval; SEM= Standard error of measurement; ANT= Anterior; PM= Posteromedial; PL= Posterolateral; COMP= Composite
*Day 1 (n=25) and Day 2 (n=21)

Means, standard deviations, medians, normalized mean reach distances, and ICC values were calculated for "modified" YBT scores from day 1 and 2 and are expressed in Table 4. Interrater reliability of "modified" YBT scores from day 1 were excellent for all reach directions and COMP scores of the right limb (ICC 0.973-0.992) and left limb (ICC 0.999-1.000) except for the left ANT reach which was good (ICC 0.860). Interrater reliability of "modified" YBT scores on day 2 were excellent for the all reach directions and COMP scores of the right limb and left limb (ICC 0.998-0.999)

## Test-retest (intrarater) reliability

Test-retest (intrarater) reliability, measures of error, and MDC scores are represented in Table 5. Test-retest (intrarater) reliability of YBT scores were moderate to excellent right limb (ICC PM 0.681-ANT 0.908) and moderate to good for left limb (ICC PL 0.714 - ANT 0.811). Less than 10% between

**Table 4.** *Modified‡ YBT reach distance and interrater reliability.*

| | Mean reach distance in cm (SD) | | Normalized mean reach† (SD) | | ICC $_{(2,1)}$ (95% CI) | |
|---|---|---|---|---|---|---|
| | Day 1* | Day 2* | Day 1* | Day 2* | Day 1* | Day 2* |
| **Right Limb** | | | | | | |
| ANT | 57.2 (6.5) | 58.0 (6.7) | 70.2 (7.1) | 70.7 (7.2) | 0.998 (0.995, 0.999) | 0.999 (0.997, 1.000) |
| PM | 85.4 (9.5) | 85.4 (8.7) | 104.7 (10.0) | 104.3 (10.4) | 0.999 (0.999, 1.000) | 0.999 (0.999, 1.000) |
| PL | 84.3 (8.4) | 84.9 (7.4) | 103.5 (10.4) | 103.7 (8.7) | 0.999 (0.998, 1.000) | 0.998 (0.995, 0.999) |
| COMP | 226.9 (21.5) | 231.7 (22.0) | 92.8 (7.8) | 92.9 (7.3) | 0.999 (0.999, 1.000) | 0.999 (0.998, 1.000) |
| **Left Limb** | | | | | | |
| ANT | 58.2 (6.2) | 58.1 (6.6) | 71.3 (6.0) | 70.8 (6.6) | 0.860 (0.711, 0.935) | 0.999 (0.997, 1.000) |
| PM | 87.4 (8.2) | 88.0 (8.7) | 107.3 (9.3) | 107.4 (8.9) | 1.000 (0.999, 1.000) | 0.998 (0.995, 0.999) |
| PL | 84.9 (7.4) | 86.5 (8.6) | 105.7 (11.4) | 105.6 (9.2) | 0.999 (0.998, 1.000) | 0.999 (0.999, 1.000) |
| COMP | 231.7 (22.0) | 232.5 (21.4) | 94.7 (7.6) | 94.6 (7.0) | 0.988 (0.973, 0.995) | 0.999 (0.998, 1.000) |

Cm= centimeters; SD= Standard Deviation; ICC= Intraclass correlation coefficient; CI= Confidence Interval; ANT= Anterior; PM= Posteromedial; PL= Posterolateral; COMP= Composite
‡Modified YBT scores were calculated from the maximum reach distance from both successful and unsuccessful trials, disregarding rater subjective decisions pertaining to trial success
†Normalized to limb length expressed as a percentage
*Day 1 (n=25) and Day 2 (n=21)

**Table 5.** *YBT reach distance test-retest reliability, error and smallest detectable change values*.*

| | ICC $_{(3,1)}$ (95% CI) | Between session total error in cm (% mean reach distance) | SEM in cm (% mean reach distance) | MDC in cm (% mean reach distance) |
|---|---|---|---|---|
| **Right Limb** | | | | |
| ANT | 0.908 (0.830, 0.950) | 2.71 (4.73%) | 0.42 (0.73%) | 1.16 (2.02%) |
| PM | 0.681 (0.479, 0.814) | 7.16 (8.47%) | 1.11 (1.31%) | 3.06 (3.62%) |
| PL | 0.756 (0.590, 0.861) | 5.42 (6.45%) | 0.84 (1.00%) | 2.32 (2.76%) |
| COMP | 0.810 (0.672, 0.893) | 12.21 (5.40%) | 1.88 (0.83%) | 5.22 (2.31%) |
| **Left Limb** | | | | |
| ANT | 0.811 (0.677, 0.894) | 3.72 (6.48%) | 0.57 (1.00%) | 1.59 (2.77%) |
| PM | 0.764 (0.544, 0.877) | 5.59 (6.47%) | 0.86 (1.00%) | 2.39 (2.77%) |
| PL | 0.714 (0.529, 0.835) | 7.28 (8.48%) | 1.12 (1.31%) | 3.11 (3.63%) |
| COMP | 0.793 (0.639, 0.885) | 13.82 (6.01%) | 2.13 (0.93%) | 5.91 (2.57%) |

ICC= Intraclass correlation coefficient; SEM= Standard error of measurement; MDC= Minimal detectable change; cm= centimeters; ANT= Anterior; PM= Posteromedial; PL= Posterolateral; COMP= Composite
*Between session measures (42 data points)

session total error was observed for all reach directions (4.73%-8.48%). SEM percentages were all less than 2% of the respective mean reach distances. Test-retest (intrarater) MDC values for the right limb ranged between 2.02% (ANT) and 3.62% (PM) and 2.31% for the COMP score. Test-retest (intrarater) MDC values for the left leg ranged from 2.77% (ANT and PM) to 3.63% (PL) and 2.57% for the COMP score.

## DISCUSSION

Reliable clinical measures of dynamic postural control may help identify individuals at higher risk of injury or assist clinicians in determining degree of functional improvement following injury. This study demonstrates that the YBT is a reliable tool for the early adolescent female population, even when administered by novice raters without significant training or experience. The YBT was developed as a modification to the Star Excursion Balance Test (SEBT) originally described by Gray.[29] The SEBT assesses reaches in eight different directions, standardizes stance limb heel position, and utilizes a testing grid marked on the ground.[30] To improve reliability and facilitate test administration, the YBT limits the amount of reaches to the three most pertinent directions, does not require the rater to simultaneously monitor stance limb heel position and reach distance, and utilizes a standardized testing device to assist with reach measurement.[31] Though studies have been performed on the reliability of the SEBT in adolescents,[32] the kinematics and muscular demands differed between the SEBT and YBT and therefore should be considered independent assessments.[33]

The findings of the current study are in agreement with other studies examining the reliability of YBT in healthy adults,[22] male collegiate soccer players,[23] and primary school-aged pre-adolescents.[24] Interrater reliability of our study ranged from ICC 0.90-0.99 which is consistent with Plisky et al (ICC 0.97-1.0)[23] where testing procedure utilized the commercially available YBT Testing Kit™. However, van Lieshout et al utilized grid marks on the floor and could explain their slightly lower interrater reliability values (ICC 0.87-0.92).[22] Additionally, measurement error in this study was mostly attributed to discrepancies in rater decisions of test success, as the "modified" YBT scores, that did not account for subjective determinations of trial acceptability, demonstrated nearly perfect interrater reliability values.

Similar to previous findings,[12] test-retest (intrarater) was found to be lower than interrater reliability and can be explained by the added variability of subject performance across testing sessions. Test-retest (intrarater) reliability of the current study ranged from 0.681-0.908 with the right PM reach score demonstrating the least between session reliability and the right ANT reach score being the most stable. These values are significantly lower than those reported by Plisky et al[23] in a population of healthy adults with a mean of 19.7 years old and a test-retest time span of 20 minutes. The sample in the current study was significantly younger, with a mean age of 12.8 years. The adolescent female is a unique population, where dynamic balance deficits are escalated by the interaction of an immature neuromuscular system, peak maturational growth rates, and emergence of sex specific differences. In studies assessing performance of unilateral balance tasks of younger populations,[24,32] between session reliability measures were also significantly lower than those of adults and further highlights population specific performance variation. Specifically, in a group of healthy adolescents from ages 12-16 years old with a test-retest interval of six days, the SEBT was found to have moderate to excellent interrater (ICC 0.59-0.95) and intrarater (ICC 0.68-0.95) reliability.[32]

The time span between testing sessions in the current study was longer (32.3 $\pm$ 9.6 days) than previous studies in an attempt to improve external validity by mimicking a typical time interval between reassessments often performed in a clinical setting. Typically, methodologies in YBT reliability studies compare two separate tests ranging from 20 minutes[23] to 7-10 days[24] apart. Though results indicated moderate to excellent reliability, the time between testing sessions may partially explain why the between session reliability values were slightly lower than previous reports[22-25] and should be considered when interpreting YBT re-assessments in early adolescent females. Despite slight reliability differences, total error and MDC values in limb reaches were similar between limbs and relative to previous work on the YBT.[22,24] In the current study, the COMP scores were least sensitive to changes between sessions demonstrating a 5.2% and 5.7% right and left leg normalized score difference, respectively. Interestingly, the ANT reach direction demonstrated the most stable between session measure, requiring the least change in value to reflect true differences in YBT performance. This is helpful as the ANT reach has been able to identify those with a greater risk of lower extremity injury[18,34] and identify those with chronic ankle instability,[35] and is often utilized as a sports pre-participation screen.

The relative measures of between session total error were similar between right and left limbs, and are slightly elevated from those reported in healthy preadolescents. Interestingly, younger preadolescents in Grades 1 and 2 performed with greater error when compared to those in Grades 3-5, and was attributed to somatotype changes during early childhood.[24] Early adolescents undergo similar changes including the redistribution of adipose tissue, increased muscular development, and leg length changes, however at more elevated rates.[2,36] The fact that this sample had a significant increase in height and weight during the one month time interval is a testament to the rapid skeletal growth at this age. Therefore, it is not surprising that dynamic postural control variations are pronounced during early childhood, plateau in later childhood, and re-emerge during early adolescence.

This study is not without limitations. Despite a sample size similar to other dynamic balance reliability studies,[23,33,37] the small sample size may not accurately reflect the true normative reach scores and MDC values for this age population. Secondly, chronological age was used as a marker of early adolescence instead of more accurate means of assessment, such as Tanner staging, onset of menarche, or radiographs. This may limit results as determination of physical maturation based on chronological age may be highly variable within this population. The sample also included a healthy population of pre-adolescents, and thus findings may not correspond to early adolescents after injury. Lastly, though the raters had limited experience in YBT administration, they were all second-year doctor of physical therapy students, and had a basic knowledge regarding movement assessment. Thus, extrapolation to other professionals involved in supervising youth athletes, such as coaches, strength and conditioning specialists, and physical education teachers, should be done with caution.

Larger cohort research is needed in the future to better quantify normative values in this high-risk population. Additionally, future studies regarding the role of the YBT in injury screening is warranted to mitigate the negative effects of the maturational changes in females prevalent during this developmental phase. Finally, research pertaining to the responsiveness of the YBT following implementation of targeted neuromuscular training programs in those high-risk individuals is needed to determine the tool's utility in the prevention and management of lower extremity injury.

## CONCLUSION

The results of this study indicate that the YBT is a reliable tool that can be utilized to identify limb asymmetries and dynamic balance deficits in early adolescent females. Differences between raters were mainly attributed to subjective determinations regarding test success and not quantitative measurement of reach distance. Additionally, the YBT is a stable measure of neuromuscular control within this population over one month, and thus may be utilized within a battery of clinical tests to monitor function after targeted interventions and help guide decision-making regarding activity progression.

## REFERENCES


1. Ott SM. Attainment of peak bone mass. *J Clin Endocrinol Metab.* 1990;71(5):1082A-1082C.
2. Lee AJ, Lin WH. The influence of gender and somatotype on single-leg upright standing postural stability in children. *J Appl Biomech.* 2007;23(3):173-179.
3. Beunen G, Malina RM. Growth and physical performance relative to the timing of the adolescent spurt. *Exerc Sport Sci Rev.* 1988;16:503-540.
4. Quatman CE, Ford KR, Myer GD, Hewett TE. Maturation leads to gender differences in landing force and vertical jump performance: a longitudinal study. *Am J Sports Med.* 2006;34(5):806-813.
5. Abderson SaH, SS, eds. *Care of the Young Athlete.* 2nd ed. Elk Grove Village, IL: American Academy of Pediatrics; 2010.
6. Tammelin T, Nayha S, Hills AP, Jarvelin MR. Adolescent participation in sports and adult physical activity. *Am J Prev Med.* 2003;24(1):22-28.
7. Steinberg N, Eliakim A, Zaav A, et al. Postural balance following aerobic fatigue tests: a longitudinal study among young athletes. *J Mot Behav.* 2016;48(4):332-340.
8. Brophy RH, Staples JR, Motley J, Blalock R, Steger-May K, Halstead M. Young females exhibit decreased coronal plane postural stability compared to young males. *HSS J.* 2016;12(1):26-31.
9. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med.* 2005;33(4):492-501.
10. Noyes FR, Barber-Westin SD, Fleckenstein C, Walsh C, West J. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. *Am J Sports Med.* 2005;33(2):197-207.
11. Arendt EA, Agel J, Dick R. Anterior cruciate ligament injury patterns among collegiate men and women. *J Athl Train.* 1999;34(2):86-92.
12. Ireland ML. The female ACL: why is it more prone to injury? *Orthop Clin North Am.* 2002;33(4):637-651.
13. Myer GD, Ford KR, Hewett TE. Rationale and clinical techniques for anterior cruciate ligament injury prevention among female athletes. *J Athl Train.* 2004;39(4):352-364.

14. Hopper AJ, Haff EE, Joyce C, Lloyd RS, Haff GG. Neuromuscular training improves lower extremity biomechanics associated with knee injury during landing in 11-13 year old female netball athletes: A randomized control study. *Front Physiol.* 2017;8:883.
15. Beck NA, Lawrence JTR, Nordin JD, DeFor TA, Tompkins M. ACL tears in school-aged children and adolescents over 20 years. *Pediatrics.* 2017;139(3).
16. Myer GD, Faigenbaum AD, Ford KR, Best TM, Bergeron MF, Hewett TE. When to initiate integrative neuromuscular training to reduce sports-related injuries and enhance health in youth? *Curr Sports Med Rep.* 2011;10(3):155-166.
17. Geldhof E, Cardon G, De Bourdeaudhuij I, et al. Static and dynamic standing balance: test-retest reliability and reference values in 9 to 10 year old children. *Eur J Pediatr.* 2006;165(11):779-786.
18. Smith CA, Chimera NJ, Warren M. Association of y balance test reach asymmetry and injury in division I athletes. *Med Sci Sports Exerc.* 2015;47(1):136-141.
19. Mayer SW, Queen RM, Taylor D, et al. Functional testing differences in anterior cruciate ligament reconstruction patients released versus not released to return to sport. *Am J Sports Med.* 2015;43(7):1648-1655.
20. Boyle MJ, Butler RJ, Queen RM. Functional movement competency and dynamic balance after anterior cruciate ligament reconstruction in adolescent patients. *J Pediatr Orthop.* 2016;36(1):36-41.
21. Garrison JC, Bothwell JM, Wolf G, Aryal S, Thigpen CA. Y balance test anterior reach symmetry at three months is related to single leg functional performance at time of return to sports following anterior cruciate ligament reconstruction. *Int J Sports Phys Ther.* 2015;10(5):602-611.
22. van Lieshout R, Reijneveld EA, van den Berg SM, et al. Reproducibility of the modified star excursion balance test composite and specific reach direction scores. *Int J Sports Phys Ther.* 2016;11(3):356-365.
23. Plisky PJ, Gorman PP, Butler RJ, Kiesel KB, Underwood FB, Elkins B. The reliability of an instrumented device for measuring components of the star excursion balance test. *N Am J Sports Phys Ther.* 2009;4(2):92-99.
24. Faigenbaum AD, Myer GD, Fernandez IP, et al. Feasibility and reliability of dynamic postural control measures in children in first through fifth grades. *Int J Sports Phys Ther.* 2014;9(2):140-148.
25. Shaffer SW, Teyhen DS, Lorenson CL, et al. Y-balance test: a reliability study involving multiple raters. *Mil Med.* 2013;178(11):1264-1270.
26. Hertel JMSD, CR. Intratester and intertester reliability during the star excursion balance test. *J Sport Rehabil.* 2000;9(2):104-116.
27. Kottner J, Gajewski BJ, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS). *Int J Nurs Stud.* 2011;48(6):659-660.
28. Fisher RA. *Statistical methods for research workers.* 12th ed. Edinburgh,: Oliver and Boyd; 1954.
29. Gray G. Lower Extremity Functional Profile. Adrian, MI: Wynn Marketing, INc; 1995.
30. Gribble PA, Hertel J, Plisky P. Using the star excursion balance test to assess dynamic postural-control deficits and outcomes in lower extremity injury: a literature and systematic review. *J Athl Train.* 2012;47(3):339-357.
31. Hertel J, Braham RA, Hale SA, Olmsted-Kramer LC. Simplifying the star excursion balance test: analyses of subjects with and without chronic ankle instability. *J Orthop Sports Phys Ther.* 2006;36(3):131-137.
32. Shaikh AW. Reliability of the star excursion balance test (SEBT) in healthy children of 12-16 Years. *Indian J of Physiotherapy & Occupational Therapy.* 2014;8(2):29-32.
33. Fullam K, Caulfield B, Coughlan GF, Delahunt E. Kinematic analysis of selected reach directions of the star excursion balance test compared with the Y-balance test. *J Sport Rehabil.* 2014;23(1):27-35.
34. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star excursion balance test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther.* 2006;36(12):911-919.
35. Hubbard TJ, Kramer LC, Denegar CR, Hertel J. Correlations among multiple measures of functional and mechanical instability in subjects with chronic ankle instability. *J Athl Train.* 2007;42(3):361-366.
36. Malina R, Bouchard, C. Bar-Or, O. . *Growth, Maturation and Physical Activity.* 2nd ed. Champaign, IL: Human Kinetics; 2004.
37. Gribble PA, Kelly SE, Refshauge KM, Hiller CE. Interrater reliability of the star excursion balance test. *J Athl Train.* 2013;48(5):621-626.